Aplicación de la sindicación de contenidos a la gestión de catálogos bibliográficos

Manuel BLÁZQUEZ OCHANDO
Juan Antonio MARTÍNEZ COMECHE
Facultad de Ciencias de la Documentación de la UCM

Recibido: 21-1-2010
Aceptado: 25-2-2010

RESUMEN
Se estudia la aplicación de la técnica de sindicación a la difusión de colecciones bibliográficas de carácter textual. Tras preparar colecciones de diversos tamaños en los formatos ATOM, RSS 1.0, RSS 2.0, MARC-XML abreviado y extendido, se comparan los tiempos de los procesos de conversión, exportación e importación. El formato que se muestra más apto es MARC-XML aunque no es un formato de sindicación propiamente dicho. Las pruebas muestran que con colecciones inferiores a 25000 registros se obtiene siempre un tiempo de importación inferior al minuto. Se concluye que la sindicación es una técnica útil para la transmisión de catálogos bibliográficos, siendo una alternativa al protocolo Z39.50, más complejo y difícil de usar.

Palabras-clave: Sindicación de contenidos, catálogos bibliográficos, gestión bibliográfica, ATOM, RSS, MARC-XML

Application of syndication to the management of bibliographic catalogs

ABSTRACT
The application of syndication to management of textual bibliographic collections is analyzed. After developing collections of several size in ATOM, RSS 1.0, RSS 2.0, short and extended MARC-XML formats, times of conversion, creation, and insertion of syndicated data are compared. MARC-XML is shown to be the most complet format, although it is not properly a syndication format. Tests carried show that with collections smaller than 25,000 records, the insertion/import time is less than one minute. The analysis suggests that content syndication is a useful technique for the transmission and retrieval of textual bibliographic data, being an alternative to the use of Z39-50 protocol, more complex and difficult to use.

Key-words: Syndication contents, bibliographic catalogs, bibliographic management, ATOM, RSS, MARC-XML

1. INTRODUCCIÓN

La sindicación de contenidos es una técnica de transmisión de información en lenguaje XML (Brickley; Guha, 2004) mediante canales o fuentes que pueden ser actualizados y compartidos con cualquier cliente en red. Esta técnica de reciente creación, uno de cuyos pioneros es Dave Winer (Winer, 2001), se aplicó inicialmente a los medios de comunicación social (New York Times, 2009). Actualmente se ha extendido a diversos ámbitos, destacando el académico, especializándose en contenidos textuales y audiovisuales.

En el campo específico de la Documentación, las aplicaciones de la sindicación consisten en canales de información generales y alertas bibliográficas (ANU, 2004), la redifusión de artículos y contenidos de revistas científicas (Abadal, E.; Estivill, A.; Franganillo, J., 2006) y en la difusión selectiva de la información en bibliotecas digitales (Peis, E.; Herrera-Viedma, E.; Morales-del-Castillo, J-M., 2008).

En este artículo se analiza la aplicación de la sindicación en el campo bibliotecario, específicamente las posibilidades de difusión y recuperación de catálogos bibliográficos textuales de diferentes dimensiones mediante técnicas de sindicación de contenidos.

2. METODOLOGÍA Y RECOPILACIÓN DE DATOS

En primer lugar, nos hemos visto en la necesidad de crear las colecciones de prueba cuyo volumen y tamaño figuran en la siguiente tabla:

Colección	Tamaño en disco (MB)	Nº Registros
1000_reg	0.77	1001
5000_reg	2.68	5002
10000_reg	5.05	10004
25000_reg	13.33	25008
50000_reg	28.34	50036
100000_reg	54.95	100054
250000_reg	144.00	250146
500000_reg	280.49	500309
1000000_reg	561.39	1000039

Tabla 1. Características de las colecciones creadas

Dichas colecciones han sido desarrolladas a partir del catálogo bibliográfico de la Library of Congress [Library of Congress, 2009], mediante consultas de temática muy general.

Una vez formadas las colecciones en formato CSV, mediante un programa (desarrollado en PHP) se obtuvieron las colecciones en XML crudo, agrupadas de mil en mil registros con el objetivo de mejorar la velocidad de creación de las tablas de datos en MySQL. Este proceso lo hemos denominado conversión de datos. En la siguiente tabla se resumen los tiempos de conversión obtenidos:

Colección	Tiempo de conversión (Segundos)
1000_reg	1.21
5000_reg	4.93
10000_reg	9.45
25000_reg	24.54
50000_reg	50.11
100000_reg	99.13
250000_reg	251.91
500000_reg	504.23
1000000_reg	992.36

Tabla 2. Tiempos de conversión de formato CSV a XML crudo

A continuación se procedió a sindicar las colecciones en diferentes formatos (ATOM, RSS 1.0, RSS 2.0), con el objeto de comparar el rendimiento de los distintos formatos utilizados actualmente en la sindicación, además de incluir en las pruebas otro formato (MARC-XML con dos variantes: extendido y abreviado) no empleado actualmente como formato propio de la sindicación, aunque si ampliamente utilizado en el mundo bibliotecario.

Dentro de las posibilidades de cada formato de sindicación, se han escogido todas aquellas etiquetas que, siendo las habituales en la práctica, sean útiles para describir registros bibliográficos textuales (no multimedia), evitando en la medida de lo posible pérdidas de información. En el caso concreto de los formatos MARC, se ha considerado un corpus bibliográfico formado mayoritariamente por monografías y se ha empleado la Clasificación Decimal Dewey, dado que la colección fuente procede de la Library of Congress.

La estructura elegida para el formato ATOM (Atom Syndication Format, 2005) es la siguiente:

ATOM

```
<entry>
<id>Número de identificación</id>
<title>Área de título</title>
<author><name>Área de mención de responsabilidad</name></author>
<updated>Fecha de actualización del registro bibliográfico</updated>
<content>Registro bibliográfico completo</content>
<link rel='alternate' href= 'URL permanente del registro bibliográfico'/>
<summary>Resumen de contenido</summary>
<category term='Temática del documento, mediante palabras clave o clasificaciones'/>
<contributor><name>Otras menciones de responsabilidad</name></contributor>
<published>Área de publicación</published>
<source>URL del canal origen del registro</source>
<rights>Derechos sobre el registro bibliográfico</rights>
</entry>
```

FIG. 1. Estructura del registro bibliográfico en formato ATOM

Las etiquetas escogidas para configurar el canal de sindicación en formato RSS 1.0 (RDF Site Summary 1.0, 2008), junto con las especificaciones de los módulos incluidos (RDF Site Summary 1.0 Modules: Dublin Core, 2000; RDF Site Summary 1.0 Modules: Content, 2002; RDF Site Summary 1.0 Modules: Prism, 2002; RDF Site Summary 1.0 Modules: Skos, 2009) son las siguientes:

RSS 1.0

```
<item>
<dc:title>Área de título</dc:title>
<dc:creator>Área de mención de responsabilidad</dc:creator>
<dc:contributor>Otras menciones de responsabilidad</dc:contributor>
<dc:publisher>Área de publicación</dc:publisher>
<dc:date>Fecha de publicación</dc:date>
<dc:type>Tipo de contenido (texto, imagen, sonido) y género (conforme a un vocabulario previo,
por ejemplo: monografías, publicaciones periódicas)</dc:type>
<dc:format>Denominación del formato del documento original, tamaño o duración</dc:format>
<dc:identifier>Identificador unívoco del documento o URL permanente del registro bibliográfico
</dc:identifier>
<dc:subject>Temática del documento, mediante palabras clave o clasificaciones</dc:subject>
<dc:source>URL del canal origen del registro</dc:source>
<dc:language>Idioma del documento</dc:language>
<dc:relation>Contenidos relacionados (por ejemplo: colección a la que pertenece el documento,
otras fuentes y recursos relacionados)</dc:relation>
<dc:coverage>Cobertura espacial o temporal del contenido del documento (por ejemplo siglo
XIX, España)</dc:coverage>
<dc:rights>Derechos sobre el registro bibliográfico</dc:rights>
<prism:publicationName>Denominación de la publicación periódica</prism:publicationName>
```

21	**<prism:edition>**Área de edición de la revista**</prism:edition>**
22	**<prism:publisher>**Editor de la revista**</prism:publisher>**
23	**<prism:publicationDate>**Fecha de publicación del documento**</prism:publicationDate>**
24	**<prism:issn>**ISSN**</prism:issn>**
25	**<skos:note>**Nota de contenido**</skos:note>**
26	**<content:encoded>**Registro bibliográfico completo**</content:encoded>**
27	**</item>**

FIG. 2. Estructura del registro bibliográfico en formato RSS 1.0

A partir de las especificaciones del formato RSS 2.0 (RSS 2.0 Specification, 2003), se ha configurado el registro de la siguiente manera:

RSS 2.0

1	**<item>**
2	**<title>**Área de título**</title>**
3	**<author>**Área de mención de responsabilidad**</author>**
4	**<pubDate>**Fecha de publicación**</pubDate>**
5	**<source>**Fuente de procedencia del item**</source>**
6	**<guid>**Identificador unívoco del item**</guid>**
7	**<link>**URL permanente del registro bibliográfico**</link>**
8	**<description>**Resumen de contenido**</description>**
9	**<enclosure>**URL, longitud y tipo MIME del documento original**</enclosure>**
10	**<comments>**URL de la página de comentarios y valoraciones del documento
11	original**</comments>**
12	**<category>**Temática del documento, mediante palabras clave o clasificaciones**</category>**
13	**</item>**

FIG. 3. Estructura del registro bibliográfico en formato RSS 2.0

Para la configuración de los registros en MARC-XML se ha partido de las especificaciones de la Library of Congress y de la MARC Standards Office (MARC-XML Schema, 2009). En las figuras 4 y 5 se recogen respectivamente las estructuras de las versiones abreviada y extendida):

MARC 1 (abreviado)

1	**<record>**
2	
3	**<controlfield tag='001'>**Número de Control Interno (Por ejemplo Registro o código de
4	barras)**</controlfield>**
5	**<controlfield tag='003'>**Número de Control para la identificación del
6	documento**</controlfield>**
7	
8	**<datafield tag='017' ind1='' ind2=''>**
9	**<subfield code='a'>**Depósito Legal o Copyright**</subfield>**
10	**</datafield>**

```
11
12 <datafield tag='020' ind1='' ind2=''>
13 <subfield code='a'>ISBN</subfield>
14 </datafield>
15
16 <datafield tag='022' ind1='0' ind2=''>
17 <subfield code='a'>ISSN</subfield>
18 </datafield>
19
20 <datafield tag='035' ind1='' ind2=''>
21 <subfield code='a'>Número de Control del Sistema</subfield>
22 </datafield>
23
24 <datafield tag='041' ind1='0' ind2=''>
25 <subfield code='a'>Código del idioma del documento original</subfield>
26 </datafield>
27
28 <datafield tag='043' ind1='' ind2=''>
29 <subfield code='c'>Código geográfico del documento original</subfield>
30 </datafield>
31
32 <datafield tag='082' ind1='' ind2=''>
33 <subfield code='a'>Clasificación Decimal Dewey</subfield>
34 </datafield>
35
36 <datafield tag='100' ind1='1' ind2=''>
37 <subfield code='a'>Autor personal</subfield>
38 </datafield>
39
40 <datafield tag='245' ind1='1' ind2=''>
41 <subfield code='a'>Área de título</subfield>
42 </datafield>
43 <datafield tag='250' ind1='' ind2=''>
44 <subfield code='a'>Nº de edición</subfield>
45 <subfield code='b'>Mención de edición</subfield>
46 </datafield>
47
48 <datafield tag='260' ind1='' ind2=''>
49 <subfield code='a'>Lugar de publicación</subfield>
50 <subfield code='b'>Editorial</subfield>
51 <subfield code='c'>Año de publicación</subfield>
52 </datafield>
53
54 <datafield tag='300' ind1='' ind2=''>
55 <subfield code='a'>Área de descripción física</subfield>
56 </datafield>
57
58 <datafield tag='310' ind1='' ind2=''>
59 <subfield code='a'>Periodicidad</subfield>
60 </datafield>
61
```

```
62 <datafield tag='490' ind1='0' ind2=''>
63 <subfield code='a'>Serie o colección</subfield>
64 <subfield code='v'>Nº de serie o colección</subfield>
65 </datafield>
66
67 <datafield tag='500' ind1='' ind2=''>
68 <subfield code='a'>Área de notas</subfield>
69 </datafield>
70
71 <datafield tag='654' ind1='0' ind2=''>
72 <subfield code='a'>Temática del documento, mediante palabras clave o
73 clasificaciones</subfield>
74 </datafield>
75
76 </record>
```

FIG. 4. Estructura del registro bibliográfico en formato MARC-XML abreviado

MARC 2 (extendido)

```
1  <marc:record>
2
3  <marc:controlfield tag='001'>Número de Control Interno (Por ejemplo Registro o código
4  de barras)</marc:controlfield>
5  <marc:controlfield tag='003'>Número de Control para la identificación del documento
6  </marc:controlfield>
7
8  <marc:datafield tag='017' ind1='' ind2=''>
9  <marc:subfield code='a'>Depósito Legal o Copyright</marc:subfield>
10 </marc:datafield>
11
12 <marc:datafield tag='020' ind1='' ind2=''>
13 <marc:subfield code='a'>ISBN</marc:subfield>
14 </marc:datafield>
15
16 <marc:datafield tag='022' ind1='0' ind2=''>
17 <marc:subfield code='a'>ISSN</marc:subfield>
18 </marc:datafield>
19
20 <marc:datafield tag='035' ind1='' ind2=''>
21 <marc:subfield code='a'>Número de Control del Sistema</marc:subfield>
22 </marc:datafield>
23
24 <marc:datafield tag='041' ind1='0' ind2=''>
25 <marc:subfield code='a'>Código del idioma del documento original</marc:subfield>
26 </marc:datafield>
27
28 <marc:datafield tag='043' ind1='' ind2=''>
29 <marc:subfield code='c'>Código geográfico del documento original</marc:subfield>
30 </marc:datafield>
```

```
31
32  <marc:datafield tag='082' ind1='' ind2=''>
33  <marc:subfield code='a'>Clasificación Decimal Dewey</marc:subfield>
34  </marc:datafield>
35
36  <marc:datafield tag='100' ind1='1' ind2=''>
37  <marc:subfield code='a'>Autor personal</marc:subfield>
38  </marc:datafield>
39
40  <marc:datafield tag='245' ind1='1' ind2=''>
41  <marc:subfield code='a'>Área de título</marc:subfield>
42  </marc:datafield>
43
44  <marc:datafield tag='250' ind1='' ind2=''>
45  <marc:subfield code='a'>Nº de edición</marc:subfield>
46  <marc:subfield code='b'>Mención de edición</marc:subfield>
47  </marc:datafield>
48
49  <marc:datafield tag='260' ind1='' ind2=''>
50  <marc:subfield code='a'>Lugar de publicación</marc:subfield>
51  <marc:subfield code='b'>Editorial</marc:subfield>
52  <marc:subfield code='c'>Año de publicación</marc:subfield>
53  </marc:datafield>
54
55  <marc:datafield tag='300' ind1='' ind2=''>
56  <marc:subfield code='a'>Área de descripción física</marc:subfield>
57  </marc:datafield>
58
59  <marc:datafield tag='310' ind1='' ind2=''>
60  <marc:subfield code='a'>Periodicidad</marc:subfield>
61  </marc:datafield>
62
63  <marc:datafield tag='490' ind1='0' ind2=''>
64  <marc:subfield code='a'>Serie o colección</marc:subfield>
65  <marc:subfield code='v'>Nº de serie o colección</marc:subfield>
66  </marc:datafield>
67
68  <marc:datafield tag='500' ind1='' ind2=''>
69  <marc:subfield code='a'>Área de notas</marc:subfield>
70  </marc:datafield>
71
72  <marc:datafield tag='654' ind1='0' ind2=''>
73  <marc:subfield code='a'>Temática del documento, mediante palabras clave o
74  clasificaciones</marc:subfield>
75  </marc:datafield>
76
77  </marc:record>
```

FIG. 5. Estructura del registro bibliográfico en formato MARC-XML extendido

A partir de las colecciones de la Tabla 1, un programa en PHP genera los canales de sindicación en los diferentes formatos con las estructuras anteriores. Los correspondientes tiempos de creación obtenidos son los siguientes:

Formato	Colección	Tiempo de creación (Segundos)
ATOM	1000_reg	0.19
	5000_reg	0.74
	10000_reg	1.32
	25000_reg	3.64
	50000_reg	8.52
	100000_reg	20.90
	250000_reg	90.89
	500000_reg	287.61
	1000000_reg	1032.25
RSS 1.0	1000_reg	0.17
	5000_reg	0.65
	10000_reg	1.65
	25000_reg	4.34
	50000_reg	8.80
	100000_reg	24.59
	250000_reg	100.34
	500000_reg	312.98
	1000000_reg	1068.93
RSS 2.0	1000_reg	0.11
	5000_reg	0.94
	10000_reg	1.21
	25000_reg	3.37
	50000_reg	8.13
	100000_reg	20.28
	250000_reg	89.51
	500000_reg	290.36
	1000000_reg	1036.30
MARC-XML 1 (abreviado)	1000_reg	0.24
	5000_reg	0.81
	10000_reg	1.75
	25000_reg	4.82
	50000_reg	11.78
	100000_reg	26.68
	250000_reg	105.28
	500000_reg	321.08
	1000000_reg	1095.83

MARC-XML 2 (extendido)	1000_reg	0.20
	5000_reg	0.84
	10000_reg	1.70
	25000_reg	4.50
	50000_reg	10.51
	100000_reg	24.77
	250000_reg	99.03
	500000_reg	326.96
	1000000_reg	1091.31

Tabla 3. Tiempos de creación de canales de sindicación por formato

Sindicadas todas las colecciones (desde 1000 hasta 1 millón de registros) en los distintos formatos, el proceso de difusión de los datos presentes en un canal, desde el servidor hasta el equipo cliente, es simulado por un programa de importación (igualmente desarrollado en PHP) cuyos objetivos esenciales son:

- Detectar el tipo de formato del canal a través de su extensión.
- Crear una tabla de datos en MySQL cuya estructura se amolde a la del formato de sindicación correspondiente.
- Leer secuencialmente cada registro bibliográfico. El tiempo ocupado en este proceso representa el tiempo de transferencia entre el equipo servidor y el equipo cliente. Este tiempo de transferencia depende en la práctica de muchos factores, entre los que destaca el ancho de banda de la red, la velocidad de proceso del equipo cliente o la memoria del sistema. En consecuencia, no consideraremos en este estudio el tiempo de transferencia obtenido.
- Insertar cada registro bibliográfico leído (en grupos de mil, por haberse comprobado que es el nivel de agrupación que minimiza los tiempos de inserción) en la tabla de datos de destino. El tiempo ocupado en esta operación lo hemos denominado inserción de datos. En este estudio, dado que no hemos considerado el proceso de transferencia, el tiempo de inserción coincide con el tiempo total de importación de las fuentes sindicadas.

Los tiempos obtenidos en el proceso de importación de fuentes sindicadas, considerando las distintas colecciones y los diferentes formatos, se resumen en la siguiente tabla:

Formato	Colección	Tiempo de Inserción / Importación
ATOM	1000_reg	0,75
	5000_reg	3,10
	10000_reg	5,27
	25000_reg	17,24
	50000_reg	42,35
	100000_reg	86,27
	250000_reg	284,09
	500000_reg	630,23
	1000000_reg	1832,74
RSS 1.0	1000_reg	1,05
	5000_reg	3,45
	10000_reg	6,55
	25000_reg	21,27
	50000_reg	54,73
	100000_reg	113,13
	250000_reg	351,60
	500000_reg	930,99
	1000000_reg	2229,34
RSS 2.0	1000_reg	0,45
	5000_reg	3,45
	10000_reg	5,83
	25000_reg	20,21
	50000_reg	50,41
	100000_reg	105,17
	250000_reg	363,52
	500000_reg	794,03
	1000000_reg	2519,13
MARC-XML 1 (abreviado)	1000_reg	1,68
	5000_reg	8,36

	10000_reg	16,85
	25000_reg	42,63
	50000_reg	92,88
	100000_reg	184,64
	250000_reg	510,92
	500000_reg	1034,99
	1000000_reg	2857,61
MARC-XML 2 (extendido)	1000_reg	4,05
	5000_reg	8,45
	10000_reg	17,21
	25000_reg	43,11
	50000_reg	92,56
	100000_reg	188,49
	250000_reg	508,32
	500000_reg	1125,76
	1000000_reg	2749,38

Tabla 4. Tiempo de inserción de datos

3. ANÁLISIS DE LOS DATOS

En primer lugar se observa que cada formato presenta unas posibilidades diferentes de sindicación de catálogos bibliográficos. Esta distinta adaptabilidad se debe a que poseen estructuras internas dispares, permitiendo o impidiendo la inserción de determinados tipos de información del registro bibliográfico.

Desde el punto de vista documental, todo registro bibliográfico consta de las siguientes áreas de información esenciales: Área de título y mención de responsabilidad, edición, clase de documento, publicación, descripción física, serie y notas (International Standard Bibliographic Description, 2009). A ellos añadiremos la posibilidad de incluir el registro bibliográfico completo, porque favorece la recuperación de información mediante cualquier aspecto o punto de acceso secundario que no haya sido tenido en cuenta en la estructura original del formato.

En la siguiente tabla se exponen las áreas de información que es posible incluir en cada formato:

	ATOM	RSS1	RSS2	MARC 1	MARC 2
Área de título y mención de responsabilidad	X	X	X	X	X
Área de edición				X	X
Área de clase de documento		X		X	X
Área de publicación	X	X		X	X
Área de descripción física				X	X
Área de serie				X	X
Área de notas		X		X	X
Registro bibliográfico completo	X	X	X	X	X

Tabla 5. Adaptabilidad bibliográfica de los formatos de sindicación

De la tabla se desprende que el formato más adecuado para la difusión sindicada de registros bibliográficos es la familia de formatos MARC, pues permite incluir todas las áreas básicas de descripción bibliográfica con el nivel de exhaustividad deseado. Dada la posibilidad de exhaustividad completa en la descripción bibliográfica, se entiende que no incluya la opción de añadir el registro bibliográfico completo, dado que supondría duplicar la información. Sin embargo, su presencia favorecería la visualización del registro en su integridad en caso necesario y el acceso a cualquier dato en un solo campo.

De los formatos de sindicación propiamente dichos, RSS1 es el que mejor se adapta a un registro bibliográfico tanto monográfico como seriado, debido a la posibilidad de incluir módulos de Dublin Core y PRISM. Aún así, se detectan deficiencias importantes, como la imposibilidad de incluir las áreas de edición, descripción física y serie. Estos problemas pueden ser solventados gracias a la posibilidad de incluir un campo de registro bibliográfico completo en el que añadir dichos datos.

A un nivel similar de adaptabilidad se sitúan ATOM y RSS2. Ambos presentan una capacidad baja de representación bibliográfica, básicamente limitada al área de título y mención de responsabilidad. RSS2 presenta la desventaja añadida de no incluir el registro bibliográfico completo. Este problema puede ser solucionado, aunque de manera poco ortodoxa, agregando módulos diseñados originalmente para el formato RSS1 (Dublin Core y PRISM) previa introducción de los *namespace*

correspondientes. Esta solución originaría un formato híbrido que dista bastante del original, convirtiéndose en un sucedáneo de RSS1.

El siguiente aspecto relevante consiste en los tiempos de creación de los canales sindicados, observando su relación con los tamaños de los archivos en los diferentes formatos. Para poder observar bien esta correlación, en la siguiente gráfica se resumen los tamaños de las colecciones sindicadas según formatos:

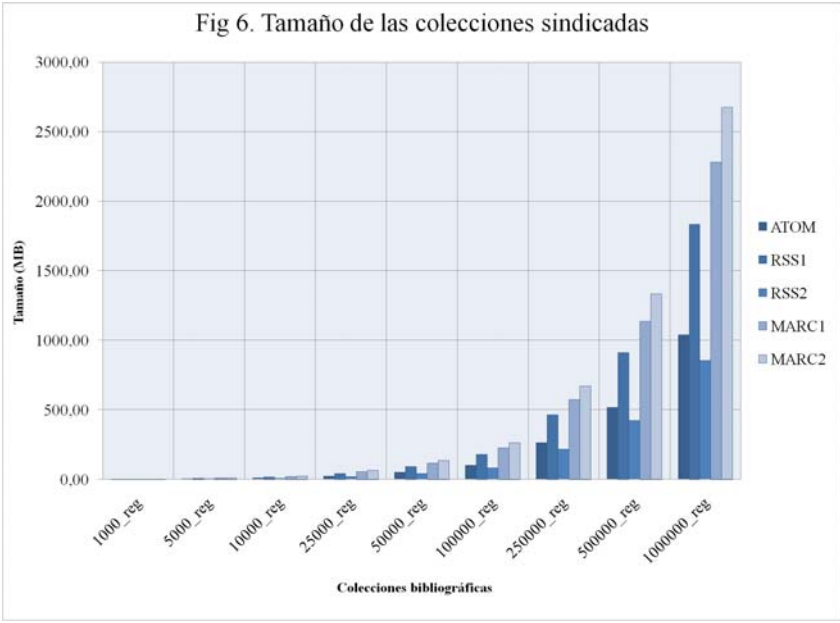

FIG. 6. Tamaño de las colecciones sindicadas

De los datos expuestos se deduce que las diferencias en la estructura de los formatos influye en el tamaño de los canales sindicados. Lógicamente cuanto más extensa y compleja es la estructura, más ocupa el canal sindicado correspondiente. Los formatos oscilan desde el más sencillo, el RSS2, hasta el más complejo, el MARC-XML extendido.

Sin embargo, las diferencias observadas en el tamaño de las colecciones sindicadas no implica unas grandes diferencias en cuanto al tiempo de creación de dichos canales, como se muestra en la siguiente gráfica:

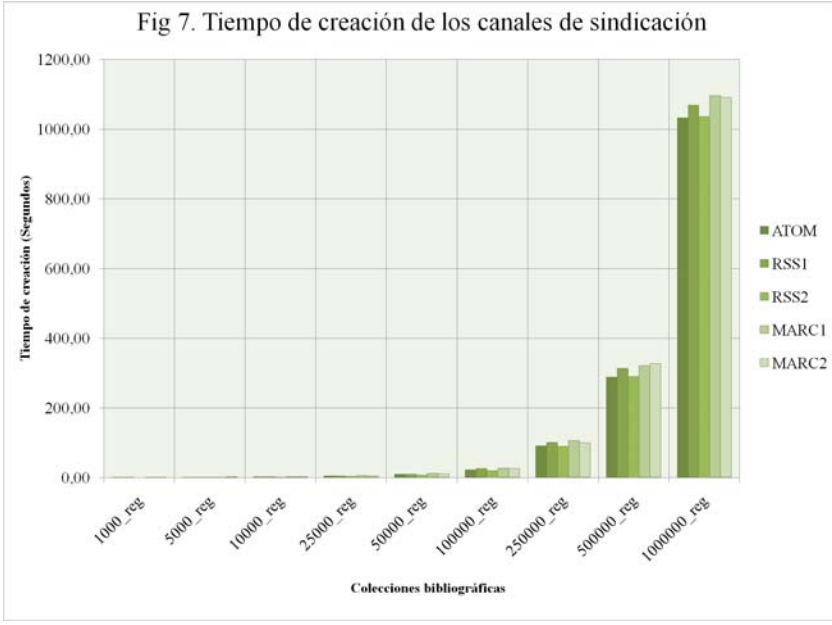

FIG. 7. Tiempo de creación de los canales de sindicación

En efecto, aunque la colección sindicada de un millón de registros bibliográficos en MARC-XML extendido ocupa 2673 MB y en RSS2 ocupa tres veces menos (854 MB), los tiempos de creación de los canales correspondientes son respectivamente de 1091 y 1036 segundos, lo que supone apenas un minuto de diferencia entre ambas. Más aún, conforme disminuye el tamaño de la colección inicial, las diferencias también disminuyen, como puede comprobarse en la Tabla 3. Por tanto, el formato (a pesar de las diferencias estructurales entre ellas) no tiene una incidencia relevante sobre el tiempo de creación de los canales sindicados correspondientes.

En cuanto a los tiempos de importación (considerando únicamente el proceso de inserción en la base de datos), los datos de la Tabla 4 se pueden resumir gráficamente de la siguiente manera:

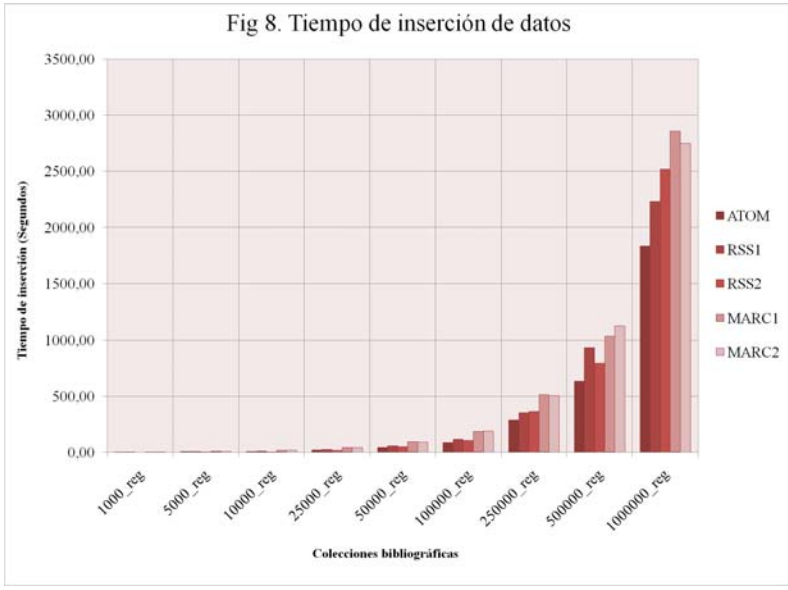

FIG. 8. Tiempo de inserción de datos

De los datos se deduce que la diferencia en el tiempo de inserción entre los distintos formatos es elevada cuando las colecciones originales son grandes, pero disminuye conforme las colecciones son menores. En el caso de las colecciones de un millón de registros, la diferencia máxima obtenida (entre los formatos MARC extendido y el formato ATOM) ronda los 15 minutos. En las colecciones de medio millón de registros, la máxima diferencia obtenida es de aproximadamente 8 minutos entre los mismos formatos. Las colecciones de 250000 registros presentan una diferencia máxima que no llega a los 4 minutos. Esta diferencia máxima es de 1.7 minutos en el caso de 100000 registros, es inferior al minuto (50.53 segundos) en el caso de 50000 registros, y disminuye hasta alcanzar los 3.3 segundos en el caso de 1000 registros.

En resumen, si bien el tiempo de creación de los canales no es relevante en relación al tamaño de las colecciones, sí lo es el tiempo de inserción/importación. De los datos analizados se deduce que la importación de datos sindicados es inferior al minuto, independientemente del formato elegido, solo cuando la colección no supera los 25000 registros, presentando unos tiempos considerables para colecciones muy grandes, superiores a los 250000 registros.

4. CONCLUSIONES

El análisis sugiere que la sindicación de contenidos es una técnica útil para la transferencia de datos bibliográficos, siendo un método alternativo al uso del protocolo Z39-50, más complejo y difícil de utilizar. Dicha utilidad es mayor cuanto menor es el tamaño de la colección sindicada, con tiempos de importación inferiores al minuto, independientemente del formato, cuando la colección consta de menos de 25000 registros. En consecuencia, esta técnica sería muy apropiada para la actualización de catálogos en bibliotecas o para el mantenimiento de grandes bases de datos bibliográficas que se nutran de múltiples fuentes.

El formato de sindicación más completo es MARC-XML, pues posee campos específicos para cualquier dato catalográfico. También RSS1 se demuestra útil y versátil para la representación de registros bibliográficos debido a la posibilidad de incluir diversos módulos de descripción especializados, como Dublin Core o PRISM, aunque carece de las áreas de edición, descripción física y serie. Sin embargo, sí posee un campo de contenidos global que permite introducir el registro bibliográfico completo, contrarrestando dichas carencias. Dicho campo global facilitaría la representación y posterior recuperación de cualquier tipo de información bibliográfica. Conforme a estos criterios, los formatos de sindicación menos aptos para la transferencia de datos bibliográficos son ATOM y RSS2.

Del análisis se desprende que no existen diferencias apreciables en cuanto a los tiempos de creación del canal de sindicación, sea cual sea el formato elegido y la complejidad de su estructura. Sin embargo, se ha comprobado que el tamaño del canal aumenta conforme aumenta la complejidad, implicando un incremento en los tiempos de importación. En consecuencia, si bien es técnicamente factible la sindicación de colecciones de cualquier tamaño, solamente con colecciones inferiores a 25000 registros nunca se alcanza un límite máximo de un minuto en la importación del canal sindicado.

5. LÍNEAS DE INVESTIGACIÓN FUTURAS

Sería interesante completar este trabajo en el futuro con la investigación del comportamiento de la sindicación de colecciones bibliográficas en entornos de red reales, para determinar fundamentalmente qué factores influyen en su rendimiento, y especialmente en el proceso de difusión de los datos.

Otro aspecto interesante de analizar consistiría en el desarrollo de técnicas de recuperación de información sobre las colecciones bibliográficas sindicadas mediante XQuery o técnicas de filtrado XPath, y su comparación con las habituales en MYSQL.

6. BIBLIOGRAFÍA

ABADAL E.; ESTIVILL. A.; FRANGANILLO. J.; GASCÓN. J. RODRÍGUEZ GAIRÍN. J.M. (2006). Sindicación de contenidos en un portal de revistas: Temaria. El Profesional de la Información. vol. 15. num. 3. p. 214-221. Disponible en: http://temaria.net/gairin2006.pdf

ANU (2004). [The Australian National Library Homepage]. Disponible en: http://anulib.anu.edu.au/lib_home.html

THE ATOM SYNDICATION FORMAT. (2005). Disponible en: http://www.atomenabled.org/developers/syndication/atom-format-spec.php (Consultado 10-03-09)

BRICKLEY. D.; GUHA. R.V. (2004). RDF Vocabulary Description Language 1.0: RDF Schema. Disponible en: http://www.w3.org/TR/rdf-schema/

International Standard Bibliographic Desciption (2009). Disponible en: http://www.ifla.org/isbd-rg (Consultado 10-03-09)

LIBRARY OF CONGRESS (2009). Library of Congress Online Catalog. Disponible en: http://catalog.loc.gov/

MARC XML Schema. (2009). Disponible en: http://www.loc.gov/standards/marcxml/ (Consultado 10-03-09)

NEW YORK TIMES (2009). The New York Times News Service/Syndicate. Disponible en: https://www.nytsyn.com/

PEIS. E.; HERRERA-VIEDMA. E.; MORALES-DEL-CASTILLO. J-M. (2008). Modelo de servicio semántico de difusión selectiva de información (DSI) para bibliotecas digitales. El Profesional de la Información. vol. 17. num. 5. p. 519-525

RDF SITE SUMMARY (RSS) 1.0. (2008). Disponible en: http://web.resource.org/rss/1.0/spec (Consultado 10-03-09)

RDF SITE SUMMARY 1.0 MODULES CONTENT ODULES: CONTENT. (2002). Disponible en: http://web.resource.org/rss/1.0/modules/content/ (Consultado 10-03-09)

RDF SITE SUMMARY 1.0 MODULES DUBLIN CORE. (2000). Disponible en: http://web.resource.org/rss/1.0/modules/dc/ (Consultado 10-03-09)

RDF SITE SUMMARY 1.0 MODULES: PRISM. (2002). Disponible en: http://www.prismstandard.org/resources/mod_prism.html (Consultado 10-03-09)

RDF SITE SUMMARY 1.0 MODULES SKOS (2009). Disponible en: http://www.w3.org/2004/02/skos/core# (Consultado 10-03-09)

RDF SITE SUMMARY.0 MODULES SYNDICATION. (2000). Disponible en: http://web.resource.org/rss/1.0/modules/syndication/ (Consultado 10-03-09)

RSS 2.0 SPECIFICATION (RSS2.0 AT HARVARD LAW). (2003). Disponible en: http://cyber.law.harvard.edu/rss/rss.html

WINER. D. (2001). Scripting News. Disponible en: http://www.scripting.com/ (Consultado 10-03-09)